Considerations on Communications Network Protocols in Deep Space[1]

Loren P. Clare
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
(818) 354-1650
Loren.P.Clare@jpl.nasa.gov

Jonathan R. Agre
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
(818) 354-9191
Jonathan.R.Agre@jpl.nasa.gov

Tsun-Yee Yan
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109
(818) 354-3016
Tsun-Yee.Yan@jpl.nasa.gov

Abstract—Communications supporting deep space missions impose numerous unique constraints that impact the architectural choices made for cost-effectiveness. We are entering the era where networks that exist in deep space are needed to support planetary exploration. Identification is made of the various distinctive elements that drive the selection of the communications protocol suite. Cost-effective performance will require a balanced integration of applicable widely used standard protocols with new and innovative designs.

TABLE OF CONTENTS

1. INTRODUCTION

Establishment of a communications infrastructure in deep space represents a fundamental advancement. As the planned the solar system exploration missions are deployed, such as the Mars exploration program, a network infrastructure to support intercommunicating entities will be implemented that operates in deep space. The question of what is the appropriate ensemble of network services that should be offered is brought into consideration. Deploying to and operating in deep space imposes many distinguishing characteristics that influence the choice of protocol suites to be used. There will be various types of communications links and traffic, and the associated protocols must cohesively integrate across interfaces. The network protocols must be able to evolve over many years and serve many missions that are still being selected.

The unique aspects of this problem domain are described together with the potential impact on the communications protocol design. These are categorized with respect to deployment, operational concerns, and the space science application domain.

The development and evolution of protocols has progressed to achieve well-established and proven solutions such as the pervasive TCP/IP protocol suite for terrestrial Internet applications. In spite of such wide acceptance and implementation, a completely universal answer still eludes us. The Internet protocol suite evolution was based on a set of underlying assumptions that includes: a) availability of a high-speed, large bandwidth backbone network, b) low propagation delay, c) low error rates, d) continuous connectivity, e) anywhere-to-anywhere traffic, f) on-demand network access, g) unlimited power and h) repairability. As these assumptions have been modified due to technological growth and new, unanticipated application requirements, enhancements to the protocol suite have become incorporated.

Several ongoing research efforts typify this situation. Modifications to TCP are being pursued for its application to geosynchronous satellite links that have high bandwidth-delay products and other characteristics requiring mitigation [1]. The Mobile IP community is investigating solutions for mobile Internet users [2,3]. The Internet Society MANET working group is developing new protocols for ad hoc networking [4]. Energy-efficient protocols are a focus of researchers in the DARPA Sensor Information Technology and Power-Aware Computing programs and related sensor network and ubiquitous computing efforts [5,6]. Quality of service capabilities for the Internet by different means (e.g., IP over ATM, Multi Protocol Label Switching) are promising realtime communications support [7,8]. Each application domain presents differing aspects that drive the need for distinct protocol features. Distinct features that arise in the deep space network context need to be considered vis-à-vis the Internet stack and other established protocols.

The primary types of deep space network links are defined as follows. Direct-To-Earth (DTE) links involve the enormous ranges from remote spacecraft (including such assets as landers on a remote planet) to Earth terminals typified by the Deep Space Network (DSN). "Proximity links"

[1] 0-7803-6599-2/01/$10.00 © 2001 IEEE
Updated November 10, 2000

are those that are contained within the general vicinity of a remote planet or other object. These proximity links may be orbiter-orbiter links, surface-orbiter links, or surface-surface links. The various link classes are illustrated for Mars in Figure 1. "Surface" links may be distinguished as being either mobile or fixed. Further, "surface" elements differ according to whether they are located on land, in the atmosphere (e.g., airplane or balloon on Mars), or subsurface (e.g., hydrobot on Europa). Additional links that will emerge in the future include Earth-orbiting elements connected to deep space. In the longer term, very long distance links such as between remote planets will be implemented, giving rise to the InterPlaNetary (IPN) Internet [9].

we have 70m diameter antenna dishes that can transmit hundreds of kilowatts of power on Earth [10], the assets at the remote ends are far more limited. This asymmetry may create a distinction in the protocol, allowing more functionality to be located in the resource-rich Earth endpoint.

Another factor in deployment is that it is much less costly to place a communications asset in orbit around a remote planet than to place it on the surface, due to the added cost of landing the element. This is exactly the opposite of the situation on Earth and may lead to placement of more functionality on the orbiter. The issue of "bent-pipe" versus sophisticated on-board functionality for Earth communications satellites has been bandied about for many years. Even today it is unclear whether the bulk of functionality should be placed in the Earth ground stations. When considered in the context of a remote planet, vastly different factors are brought to bear.

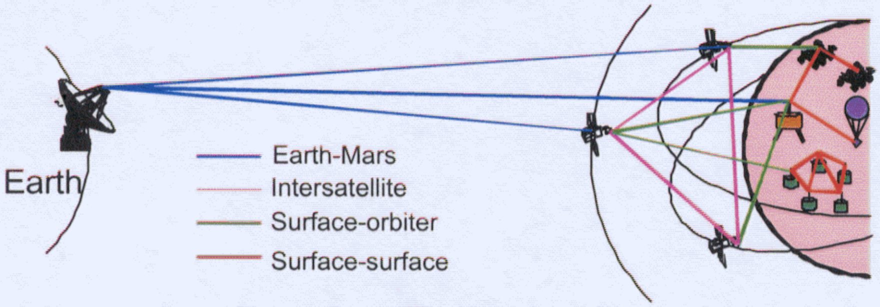


Figure 1. Types of Deep Space Links

2. Deployment Considerations

Exploration of the solar system for the next several decades will rely on robotic means. Mars orbiters and surface rovers will play a large role in the deep space data gathering missions during the near term. The establishment of a communications infrastructure for Mars will represent the genesis of deep space-based networking.

Placing assets at these very long distances implies large associated launch costs. These costs are largely driven by mass and size. All remote resources impacting communications will be limited, including power/energy, antenna size, processors and storage media. The high relative costs of physical communications (in terms of the capacity provided) will create a situation different from conventional terrestrial communications system design options. Communicating a bit across a Direct-To-Earth (DTE) link will be precious. Even proximity links in deep space will have relatively high costs. Efficiency demands on these links are typically so high that all overheads are subject to scrutiny. Despite this distinction, having the ability to cost-effectively and easily provide connectivity among a variety of applications is a major consideration. It is worthwhile to provide flexible interfaces but one must quantify the additional overhead bits that must be communicated with the underlying application information.

The deployment consideration also creates a vast difference in the physical assets at the endpoints of DTE links. While

Deployment for solar system exploration requires sophisticated navigation to reach the desired location. Navigation is highly related to the communications systems, and therefore synergy is derived by integration of the two functions. This can in turn affect the choice of the protocols used. For example, RF ranging may be achieved using the communication channel.

Remote space is desolate. This means that you must bring everything you need with you. In particular, the energy supply or the means to harvest energy *in situ* must be carried along. This creates a need to minimize the energy consumption by the remote assets. While research on energy-aware protocols is beginning for the emerging commercial mass market of mobile handheld devices (cellphones, wireless interfaces for laptops and PDAs), there are major improvements in bits/joule that can be derived for energy-sparse space applications. For example, protocols that permit the transceiver to be shut off during idle intervals can provide far more energy efficiency than, e.g., IEEE 802.11 [11], where the device is always on.

Utilization of the RF spectrum is a major concern in of all wireless protocols in office/urban environments, because there is limited spectrum to support large user communities. Spectral efficiency for maximizing the number of users in a given band and locale is of less concern for remote planet proximity links. This greater freedom may significantly impact protocol design.

Another aspect of deploying objects in deep space is that it takes a long time to place them there. For example, the Cassini spacecraft was launched October 15, 1997, and won't complete its 3.5 billion kilometer journey to Saturn until July 1, 2004. Furthermore, it is highly cost-effective to wait for specific launch window opportunities. Therefore, assets placed at a certain location (e.g., Mars) will tend to arrive at times that are spaced at rather long intervals (26 months in the case of Mars). The desire to build up a communications infrastructure will be constrained by this deployment schedule. That is, one must be able to operate during the long periods between the addition of new components and to accommodate the added capabilities. An illustrative example of a possible infrastructure buildup at Mars is depicted in Figure 2

there will be largely unknown. This implies that an "ad hoc" network will be called for, i.e., since there is no infrastructure already present and the exact placement of assets is unpredictable, the communications will have to self-organize to accommodate the situation that arises. Ad hoc multihop network technologies exist in the tactical military arena that may be leveraged for use in deep space applications. Ad hoc wireless networks currently have limited use in commercial applications. For example, cellular telephony and wireless LANs generally use single-hop wireless links to fixed access points connected to wired infrastructure. Emerging commercial standards, such as the Bluetooth "scatternet" concept [12], may present applicable solutions.

Planetary precision landings are currently not possible; the state-of-the-art may be gauged by the 100s of km error ellipse for the Mars Pathfinder. Concept missions for improving this precision include landing beacon networks [13] (recall the synergy with navigation already mentioned) are under investigation. Clearly, precision landings are prerequisite to forming robotic outposts and ultimately

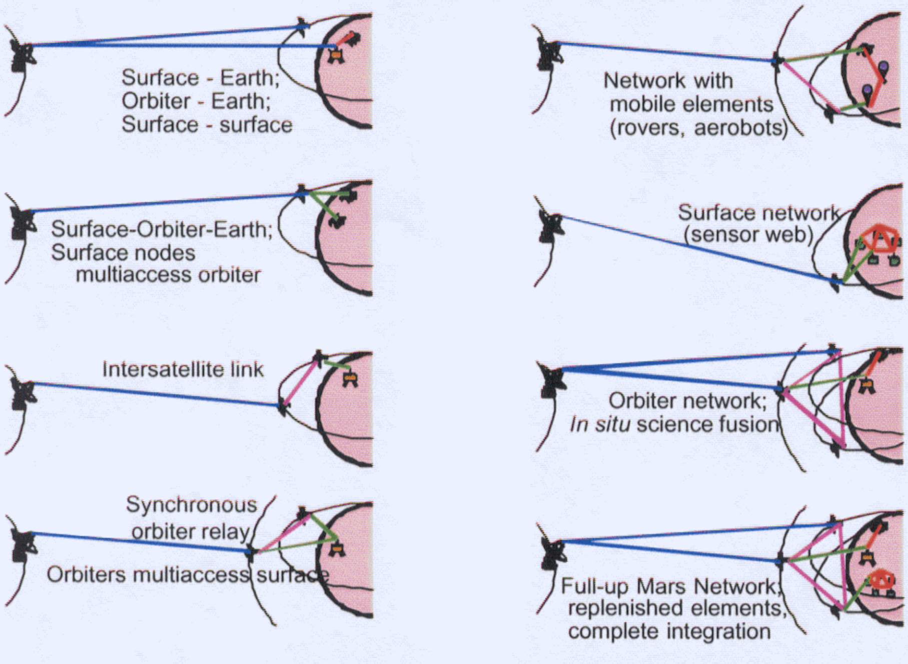


Figure 2. A possible communications infrastructure feedforward evolution.

The high cost of deploying assets in deep space will, unfortunately, limit the number of objects present. This differs from widely-used communications systems on Earth that often involve large numbers of communicating endpoints (e.g., the Internet). These commonplace protocols are designed to support this diversity and magnitude, but may do so at a cost of added overhead.

Finally, because the deep space mission is exploration, we virtually never go to the same place twice. The conditions

supporting human presence.

3. Operational Considerations

Deep space mission operations are subject the to laws of orbital mechanics. Not only does this affect launch opportunities as described above, it directly impacts the availability of communications links as orbiters orbit and surface elements revolve with the planet. Surface elements are subject to day/night operational distinctions due to solar power and coldness. These episodic links generally come and go predictably. Outages require protocols that can queue/store data while links are down, and efficiently reacquire links as they reappear. Furthermore, because communications is a precious resource in deep space, the links must be maximally utilized while they are available. For example, as an orbiter passes over a lander, the SNR will change dramatically (and largely predictably) as the slant range changes. Dynamically varying the data rate and/or the transmit power to match the SNR promises significant improvement in resource use.[2]

The large distances associated with DTE links yields very large propagation delays (e.g., approximately 10 to 20 minute one-way light time to Mars). This obviously restricts the protocol that may be used. For example, stop-and-wait ARQ error control would be absurd. The problems of using TCP window-based error/flow control associated with the "long" propagation delays of geosynchronous satellites are greatly exacerbated in DTE links. Generally, any handshaking/negotiation in the protocols must be carefully considered.

The long distances of DTE links also imply extremely small received power levels. Sophisticated physical layer solutions have been constructed. A key technology is channel coding, however, two aspects should be considered in the context of protocols. First, channel coding works best with long codes, i.e., with large chunks of data at a time. Second, if different source streams are multiplexed, it is problematic to provide distinct qualities of service (integrity).

Mobile elements such as rovers require navigation support, and often are teleoperated by Earth command in spite of the large latencies involved. The position of each surface element (e.g., in a sensor web) must be known in order to support the sensing function. In addition to the use of common assets to acquire position information, one may in turn use position knowledge to aid communications. In particular, "Mobile IP" has been offered as a solution to mobile users of the Internet, but suffers several drawbacks. If instead we may assume each asset knows its position, then this may used as its address. Further, geographic routing is achievable.

Many missions are comprised of distinct phases of operation, such as the entry/descent/landing phase versus execution of the science mission. This is distinguished from commonplace networks. For example, while certain statistical anomalies may predicted (e.g., geographic day/night "rush hours" of traffic, or the surge of telephone calls on Mother's Day), the common carrier networks generally operate in steady state. A deep space mission may gain efficiency by utilizing protocols matched to each mission phase.

As greater development of *in situ* capabilities are made, needs for intercommunications among the elements will arise. These will include realtime support operations among cooperating elements (e.g., supporting *in situ* propellant production). Protocols that maintain tight latency requirements associated with control functions are then needed. Energy efficient protocols will be essential for long mission lifetimes.

Because of the high costs, there is a strong desire to use assets across multiple missions. For example, the Mars Odyssey science orbiter that is planned to arrive at Mars in October 2001 will serve as a relay for the Mars '03 rovers. It is a stated objective that every future orbiter will provide a telecommunications relay capability, possibly extending well beyond its own scientific mission, for use by subsequent missions. This infrastructure feedforward is critically dependent on the selected protocols. It implies that the protocol must be evolvable over many years. For example, one should use a control header format that allows for expansion to support as-yet undefined services.

Missions will require realtime autonomous control, and local decision-making will be needed based on the conditions at hand. These operations will result in variability in proximity communications loads as well as connectivity of the links. Ongoing technology development of wireless self-organizing ad hoc networks can also be applied to these objectives.

As in all space missions, unpredictable faults will arise. When communicating elements fail, the network must immediately detect and recover from the resulting reduction in resources and change in topology. In addition, a substantial surge in traffic may be expected as surviving components reconfigure for degraded operations.

As was mentioned earlier, it is less expensive to place assets in orbit than on the surface of remote planets. This raises the possibility of engineering a surface mission in which the bulk of the communications assets are left in orbit, and minimal assets are placed on the surface. The surface communications would be designed for the proximity link to the orbiter, while the orbiter would have the capability to close the difficult DTE link, and store and forward data. Thus a relay architecture arises, consisting of two very different physical links. Resources (transceiver, energy, buffer capacity) corresponding to each of these links are precious and heavily utilized. If primary error control is limited to the transport layer (such as is generally the case with TCP and the Internet), inefficiencies may arise from retransmitting packets that were error-free on the first hop. (This extends further to

[2] Kar-Ming Cheung, internal JPL briefing, Nov.2, 2000.

longer multihop paths.) Thus reliable link and network layer error control protocols are more appropriate and important in the deep space domain. The CCSDS CFDP protocol provides a solution for this problem situation [14].

This architecture would support using an orbiter as a relay for multiple surface elements. The Proximity-1 link protiocol is being designed for this scenario [15]. This introduces a multiplexing/concentration function on the orbiter, so that associated queueing protocols need to be defined. If multiple surface elements simultaneously fall within the orbiter's footprint, multiaccess control must be implemented. Broadcast/multicast messaging to the surface is also a possibility. It is possible that the orbiter would relay on behalf of different surface elements using different radio (including protocol) technologies (perhaps because of being deployed years apart); in this case software radio technologies may be used to make the orbiter radio reconfigure operate as needed as it passes over each surface asset[3]. Increasing numbers of assets will introduce the need for higher level network services in the supporting architecture, such routing. However, it is likely that the number if intermediate nodes in a path will be limited, so that routing complexity may be of less concern than in terrestrial domains.

As solar system exploration evolves over time, the number of deep space entities will become increasingly large. These will result in a multitude of possible interconnections. The set of viewing windows among remote surface elements and orbiters and Earth will grow, and even though these times are predictable, they will cause the number of possible schedules that may be chosen to be immense. Another anticipated aspect is a great heterogeneity of communicating elements. Therefore, the complexity that arises may be expected to preclude straightforward optimization. Instead, the fine-grain determinism may be "ignored" in favor of protocols that operate as though these aspects are random. Control mechanisms may be implemented that enable demand access [16], thereby accommodating complexity as well as dynamically offered traffic in scheduling communications.

4. APPLICATION-SPECIFIC DRIVERS

Deep space missions, while widely variable in the scientific goals they are pursuing, nevertheless enjoy substantial commonalities from the communications network perspective. Generally, at least perhaps until missions change with significant human presence in deep space, the primary objective is exploration. This involves sensing unknown environments and providing that information to end users on Earth. Short command messages will direct the deep space assets. Advantages may be derived by tailoring the communications system for the predicted traffic patterns.

One immediately recognizable communications characteristic is that there will be a much greater volume of traffic heading toward Earth than toward deep space. Thus the DTE/Direct-From-Earth links will be highly asymmetric. Almost every conventional protocol assumes symmetry.

Because the information being delivered corresponds to sensing of heretofore unknown phenomena, the Earth scientist generally desires the maximum amount of unprocessed, raw sensor data that the communication system can support. That is, there is a fear that comparing the data with an a priori model, so that vast data reduction may be made, will corrupt the information because of incorrect assumptions in that model. Nevertheless, some local processing may performed (e.g., lossless source compression), which will then result in variability in the offered traffic rate, even if a constant sensor sampling rate were used initially. Furthermore, because of the extremely skinny DTE links, there is great benefit in increasing the scientific information per bit returned. Thus methods where the scientist can select parameters (sample rate, select different sensors, etc.) offer great benefits.

One can expect increasing scientific acceptance of autonomous operated missions. The communications network will greatly benefit from integrating *in situ* information processing for science and support operations. Greater efficiencies may be expected by performing local processing and subsequent transport of the processed information, rather than simply communicating raw high-bandwidth sensor data. These gains primarily stem from the reduced energy consumption of computation versus communications. Increased processing tends to produce outputs characterized by smaller volume (in terms of bits needed to represent them) together with greater dynamics and unpredictability. In order to accommodate this increased variability using remotely operated assets, new and adaptive communications protocols and information handling techniques are required.

An emerging class of experiments of particular interest to the communications network designer relates to "network science." This refers to measurements made simultaneously among spatially distributed sensors that capture different perspectives of the same phenomena. For example, a seismic sensor array is able to determine the absolute magnitude and to localize the epicenter of a quake, while a single sensor could not. These sensor networks may support local communication to facilitate the collection of data [17]. The volume of local communications may be much greater than that forwarded on to the long haul DTE links. Communications support for this important class of missions that is designed

[3] Jeffrey Srinivasan, JPL , private communication

specifically for the synchronization and traffic flow requirements promises great increases in efficient use of the resources.

5. Protocol Selection

It is certainly prudent to identify standardized and commonly used protocols whenever possible in designing any communications system. Such protocols are well understood, have proven success across wide applications, have broad vendor support, and user applications experience. If a conventional protocol has shortfalls in a particular application context, effort should be made to identify minimized modifications needed to meet the specific requirements.

Layered protocols, such as the Internet stack, can be modified to various degrees: the entire stack below a given layer can be replaced, a single layer can be replaced, features can be added to a layer or a layer can be restricted such that a subset of its functions is supported. It is possible to support application level transparency so that software developers can use the same application-layer function calls, such as FTP, for different lower layer protocols. For example, TCP/IP has been modified to support behavior not originally included for long delay satellite links [1]. While the operation of TCP is then different, the Applications Programming Interface (API) is the same. In IEEE 802.11 wireless LAN protocol, the problem of mobility was handled by essentially providing a networking layer function entirely within its definition, which therefore extends its scope beyond the link layer. Efforts to streamline protocols, such as TFTP [18] have eliminated or reduced functionality to improve performance.

In may expect great diversity in the links and subnetworks comprising the deep space network. In cases where the conditions call for distinct networks, gateways may be employed to connect them. The gateways will perform protocol translation; and thereby provide transparencies to the user although with greater difficulty than if the same protocol stack were used in each network. For example, a sensor network may use a special purpose energy-efficient protocol for internal communications, but employ a gateway based station to provide information and access to the users. Support for TCP/IP addressing to end-points can be provided by tunneling through the given network protocol but may do so with inefficiency.

In some instances, the conditions and requirements can be so unusual that no conventional protocol will provide adequate performance, and the cost of designing and implementing a new solution is warranted. Integration will be needed to ensure proper operation within the overall protocol stack and network architecture.

6. Conclusions

We have presented numerous aspects that are uniquely characteristic of communications in deep space. While the problems of developing solutions for physical links extending over vast distances have been under study for a number of years, we have presented newly emerging problems associated with protocols that arise as deep space-based networks are on the horizon. A discussion of the potential impact of these characteristics on the protocol design and operation has been given.

We believe that the distinct characteristics identified warrant further investigation regarding the potential cost-performance advantages that may arise in developing new or modified protocols. Perhaps a new suite of standards will apply, having greater commonality with applications such as emergency disaster relief networks or tactical unattended ground sensor networks for military.

Protocols are being designed that present the greatest promise of potential benefit when compared to use of conventional protocols. Simulation analyses are being developed to quantify the performance advantages. Ultimately, the cost savings that result from improved resource utilization must be weighed against the developmental costs to implement modified or custom-designed protocols.

References

[1] M. Allman and D. Glover, "Enhancing TCP Over Satellite Channels using Standard Mechanisms," Internet Society IETF Network Working Group RFC 2488, January 1999.

[2] C. E. Perkins, "Mobile IP," IEEE Communications Magazine, May 1997.

[3] Upkar Varshney and Ron Vetter, "Emerging Mobile and Wireless Networks," CACM, Vol. 43, No. 6, June 2000, pp.73-81.

[4] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations," RFC 2501, IETF, Internet Society, http://www.ietf.org/rfc/rfc2501.txt, January 1999.

[5] Sensor Information Technology (DARPA/ITO program), http://www.darpa.mil/ito/research/sensit, 2000.

[6] Power Aware Computing/Communication (DARPA/ITO program), http://www.darpa.mil/ito/research/pacc/index.html

[7] M. Perez et al., "ATM Signaling Support for IP over ATM," IETF RFC 1755, February 1995.

[8] Eric C. Rosen et al., 'Multiprotocol Label Switching Architecture," IETF Draft, Network Working Group, July 2000.

[9] Internet Society IPN Special Interest Group, InterPlaNetary Interne Study, http://ipn.jpl.nasa.gov/

[10] Deep Space Communications Complexes, http://deepspace.jpl.nasa.gov/dsn/brochure/complexes.html

[11] Bob O'Hara and Al Petrick, *IEEE 802.11 Handbook: A Designer's Companion*, Standards Information Network, IEEE Press, New York, NY, 1999.

[12] Bluetooth, http://www.bluetooth.net, 2000.

[13] Dan G. Tuckness, "Precision Landing on Mars Using Imaging Penetrator Beacons," J. of Spacecraft and Rockets, Vol. 31, No. 6, November-December 1994.

[14] *CCSDS File Delivery Protocol (CFDP)*. Draft Recommendation for Space Data Systems Standards, CCSDS 727.0-R-3. Red Book. Issue 3. Washington, D.C., CCSDS, June 1999.

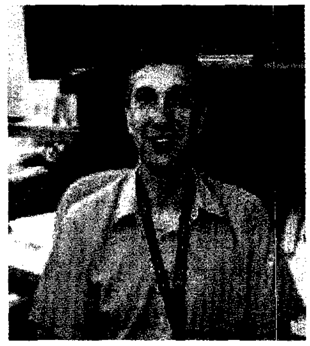
[15] *CCSDS Proximity-1 Space Link Protocol.* Draft Recommendation for Space Data Systems Standards, CCSDS 211.0-R-2. Red Book. Washington, D.C., CCSDS, January 2000.

[16] Jeffrey Hayden, "Spacecraft/Ground Architectures Using Internet Protocols to Facilitate Autonomous Mission Operations," IEEE, May 2000.

[17] Jonathan Agre and Loren Clare, "An Integrated Architecture for Cooperative Sensing Networks," IEEE Computer, Vol. 33, No. 5, May 2000, pp.106-108.

[18] K. Sollins, "The TFTP Protocol (Revision 2)," IETF RFC 1350, July 1992.

Loren Clare is the Technical Group Supervisor for the Communications Networks Research Group at the Jet Propulsion Laboratory. He obtained the Ph.D. in System Science from the University of California, Los Angeles in 1983. His research interests include wireless communications protocols, self-organizing systems, network systems design, modeling and analysis, and realtime distributed control systems. Prior to joining JPL in May 2000, Dr. Clare was a senior research scientist at the Rockwell International Science Center. Recent work focused on distributed sensor networks, including low-power protocol design as well as development of a layered architecture for cooperative signal processing. He has extensive experience in tactical network design and analysis, as well as satellite networking. He has also developed communications protocols for realtime networks supporting industrial automation.



Jonathan Agre is a Senior Member of the Technical Staff in the Communications Sysems Research Section at the Jet Propulsion Laboratory. He obtained the Ph.D. in Computer Science from the University of Maryland, College Park in 1981. His research interests include wireless sensor networks, distributed systems, communication protocols, realtime embedded systems, simulation and performance modeling. Prior to joining JPL, Dr. Agre was a senior research scientist at the Rockwell International Science Center working in the area of distributed sensor systems. He participated in many successful projects including the DARPA Low Power Integrated Microsensors (LWIM), the DARPA Adaptive Wireless Arrays for Interactive RSTA in SUO (AWAIRS), DARPA SenseIT Distributed Services for Microsensor Networks, and the ARL Sensors Fedlab Advanced Integrated Microsensors project. New technologies investigated included self-organizing protocols, power-aware algorithms, software infrastructure, distributed realtime operating systems and cooperative signal processing.

Tsun-Yee Yan is the Technical Group Supervisor for the Digital Signal Processing Research Group at the Jet Propulsion Laboratory, as well as Manager of the NASA Spacecraft Communications Technology and Applications Program. He obtained the Ph.D. from the University of California, Los Angeles in 1978. Dr. Yan has been with JPL since 1980. His research interests include high rate digital signal processing algorithms, adaptive techniques for band-limited environments, bandwidth and power efficient modulations, multiple access algorithms, optical communications, and application of artificial intelligence technology to telecommunication problems. He has led a number of successful technology development tasks for recommendations to CCSDS (Consultative Committee for Space Data Systems) and has chaired numerous technical sessions and served as technical program committee members for INFOCOM, GLOBECOM and IMSC. Dr. Yan is a former Editor of the multiple access strategies for the IEEE Transactions on Communications.